MSA101/MVE187 2017 Lecture 1

Petter Mostad

Chalmers University

September 4, 2018

CAUTION: These overheads (and those for coming classes) only contain part of the information covered in the lectures. You need to make your own notes in addition!

Frequentist issue 1: Interpretation

Example:

- We assume the numbers 4.2, 5.6 and 4.6 is a random sample from a normal distribution with expectation μ and fixed variance 1.
  As the numbers have mean 4.8, a 95% confidence interval for μ can then be computed as

$$\left[4.8 - 1.96 \cdot \frac{1}{\sqrt{3}}, 4.8 + 1.96 \cdot \frac{1}{\sqrt{3}}\right] = [3.67, 5.93]$$

- A possible interpretation: If three numbers are resampled from the distribution many times, the re-computed confidence intervals will contain μ with probability 95%.
- Another common interpretation: The interval [3.67, 5.93] contains μ with 95% probability.

What is your attitude towards misinterpretations of the confidence interval?

- People need to be better educated about the correct interpretation.
- I don't care: As long as I as a mathematician/scientist compute and present correct results, it is not my problem how it is interpreted.
- The difference between the two interpretations above is so small it is unimportant.
- Other?

Frequentist issue 2: Objectivity

Example:

- Assume we have a sequence of intepependent trials each resulting in success (1) or failure (0), with a probability of succes equal to p. Assume we have observed the following data:

  $$0, 1, 0, 0, 1, 0, 0, 1$$

  We then make the estimate $3/8 = 0.375$ for p. How "good" is this estimate?
- It is often said that an *estimator* that is unbiased is "good". Is this estimator unbiased? It depends on which estimator we have used!
- Alternative 1: The estimator is: Make 8 trials, let X be the number of successes, and compute $\hat{p} = X/8$.
- Alternative 2: The estimator is: Make trials until you have produced 3 successful trials, let X be the number of trials you needed to do, and compute $\hat{p} = 3/X$.

Continuation of example

- Exercise: Prove that the estimator in alternative 1 is unbiased (easy), and that the estimator in alternative 2 is biased (more difficult).
- Our point here: If we use the biasedness of the *estimator* to judge whether the *estimate* 0.375 is good, the result depends on which estimator we are using, which depends on what went on in the head (the plans) of the person doing the experiments.

Continuation of example

- In the same situation as above, and the same observations, we want to make a hypothesis test with $H_0 : p \geq 0.6$, and alternative hypothesis $H_1 : p < 0.6$. What is the p-value?
- To answer the question, we need to know which *test statistic* should be used.
- Alternative 1: The test statistic is: Make 8 trials and let X be the number of successes. Then, assuming $p = 0.6$, we get $X \sim \text{Binomial}(8, 0.6)$. The possible values for X and their probabilities are

0	1	2	3	4	5	6	7	8
0.001	0.008	0.041	0.124	0.232	0.279	0.209	0.090	0.017

We get that the p-value becomes 0.174; the sum of the probabilities for $X = 0, 1, 2, 3$.

Continuation of example

- Alternative 2: The test statistic is: Make trials until 3 successes have appeared and let X the number of trials necessary. Then, assuming $p = 0.6$, we get $X \sim \text{Neg-Binomial}(3, 0.6)$. The possible values for X and their probabilities are

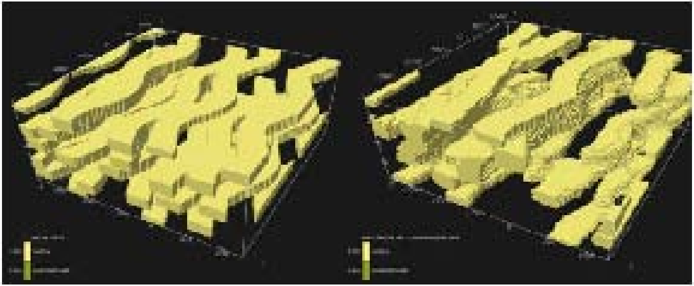
3	4	5	6	7	8	9	10	11
0.216	0.259	0.207	0.138	0.083	0.046	0.025	0.013	0.006

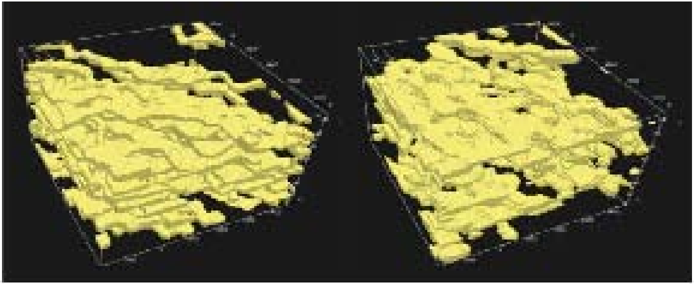
12	13	14	15	16,17,...
0.003	0.001	0.001	0.000	total 0.000

  We get the p-value 0.095; the sum of the probabilities for $8, 9, 10, \dots$.
- Note that if we use a significance level of 0.1, we will reject the null hypothesis using the second test statistic, but not using the first test statistic.

Frequentist issue 3: Repeatability

Example: Stochastic modelling of oil reservoirs





Frequentist issue 4: Contextual information

- Assume you want to find out if a coin is "fair", i.e., if the probability p for heads is 0.5. You throw the coin 8 times and get heads 2 times. What do you believe about the probability p, and how certain can you be?
- Assume you are a doctor who has received permission for a new experimental surgical procedure. After 8 procedures, 2 are successful. What do you believe about the probability p for a successful procedure, and how certain can you be?
- Assume you work at a factory and you want to make a quality control of a product. Out of 8 randomly chosen items, 2 were faulty. What do you believe about the probability p that an item is faulty, and how certain can you be?
- We saw above that what people generally want from a statistical analysis are probabilistic predictions about future observations. Generally, such predictions will need to take the context into account. If p is simply regarded as an "unknown parameter", this cannot be done.

A Bayesian approach to statistics

- Our goal is to build *stochastic models* (probabilistic models) for the real world, corresponding to our knowledge, and to use these models to make probabilistic *predictions*.
- Probability is a feature of *knowledge* of the real world, not of the real world itself.
- It is not useful to try to separate between "unknown parameters" and "random variables" in these models: All are known/unknown to some extent, and they should all be treated as random variables.
- The stochastic models are *personal* (as they model knowledge), but rational persons with the same knowledge about some part of reality should obtain the sams stochastic models for that part of reality.

Example: The future success of a surgical procedure

- Note that *predictions* are key!
- Note that predictions might vary, in that they might use different information. They are personal. There is no "correct" prediction.
- You would model the probability p for a successful procedure as a random variable, not as a parameter.
- You would model the probability density for p *without* taking into account the experiment where 2 of 8 procedures succeded: The *prior*.
- You the compute the probaility density for p *taking into account* the experment: The *posterior*.

Statistics as learning, not "estimation"

- Assume a stochastic model includes a variable X modelling some real world quantity. Assume that quantity is observed to have the value x. Then our *updated model* should be the stochastic model *conditioned on* the information $X = x$.
- Technically, this conditioning will correspond to using Bayes theorem, which is why this is called Bayesian statistics.
- In fact, all scientific learning is based on making observations. If a scientific theory is represented as a stochastic model, the process of scientific learning can be represented, to a certain approximation, as a Bayesian update of this model.

Example: Stochastic modelling of oil reservoirs

- The variable of interest might be the amount of oil, the data might be geological observations along a trial well. Many other variables describe the geological geometry.
- Not useful to estimate "parameters" from data: Knowledge about geological geometry will only increase somewhat with this particular data.
- Important to take the residual uncertainty in parameters into consideration, when predicting!

Frequentist vs Bayesian statistics

- The frequentist and Bayesian paradigms, when used on the same problem, often yield similar or identical practical results. Why?
- The two methods should share the same likelihood model. A frequentist approach for estimation followed by prediction in many cases correspond computationally to a particular choice of prior distribution on the parameters. When this prior corresponds to the one used in the Bayesian analysis, the two approaches give identical results.
- Example: Learning about a proportion p from repeated experiments. A uniform prior on $[0, 1]$ yields Bayesian results corresponding to classical ones.

Example: Intervals for expectations of normal distributions

- We assume data $x_1, \ldots, x_n$ is a random sample from a normal distribution with expectation μ and known variance $\sigma^2 = 1$.
- A frequentist analysis can compute from $x_1, \ldots, x_n$ a 95% confidence interval, say $[0.42, 0.73]$, for μ.
- People tend to interpret this as $P(0.42 \leq \mu \leq 0.73) = 0.95$. (This interpretation is wrong).
- However, if we assume a *flat prior* for μ and do a Bayesian analysis, we derive at the 95% *credibility interval* $[0.42, 0.73]$. The correct interpretation of this is exactly $P(0.42 \leq \mu \leq 0.73) = 0.95$.
- Note: We here expand the set of probability distributions to include also *improper distributions*, i.e., those that integrate (or sum) to ∞.
- For many, but not all, applications, a flat prior is reasonable.

The Bayesian paradigm for statistics

- A set of variables (discrete and/or continuous) are chosen to represent or describe some part of the real world, including
  - Variables representing observed quantities.
  - Variables representing things you want to know or predict.
  - Ancillary variables.
- A function over all possible combinations of values of the variables is established, representing the joint probability distribution.
- Some of the variables are observed (i.e., fixed) and the probability model conditional on fixing these variables is found.
- Predictions for observable quantities are made from the conditional model.